

```
<html>
<head>
<meta http-equiv="Content-Type" content="text/html; charset=iso-8859-1">
</head>
<body>
<p>Hey, diddle, diddle, the cat and the fiddle.</p>
<p>Starlight, starbright, first star I see tonight,<br>
I wish I may, I wish I might, have the wish I wish tonight.<br>
</p>
<p>The quick brown fox slyly jumped over the lazy dog.</p>
</body>
</html>
```

Figure 1

```
<html>
<head>
<meta http-equiv="Content-Type" content="text/html; charset=iso-8859-1">
</head>
<body>
<p>Mary had a little lamb, its fleece was white as snow.</p>
<p>Hey, diddle, diddle, the cat and the fiddle.</p>
<p>Starlight, starbright, first star I see tonight,<br>
I wish I may, I wish I might, have the wish I wish tonight.<br>
</p>
<p>The quick brown fox slyly jumped over the lazy dog.</p>
</body>
</html>
```

Figure 2

1000

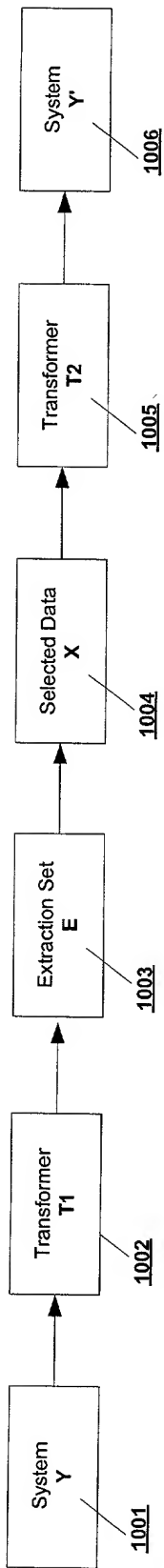


Figure 3

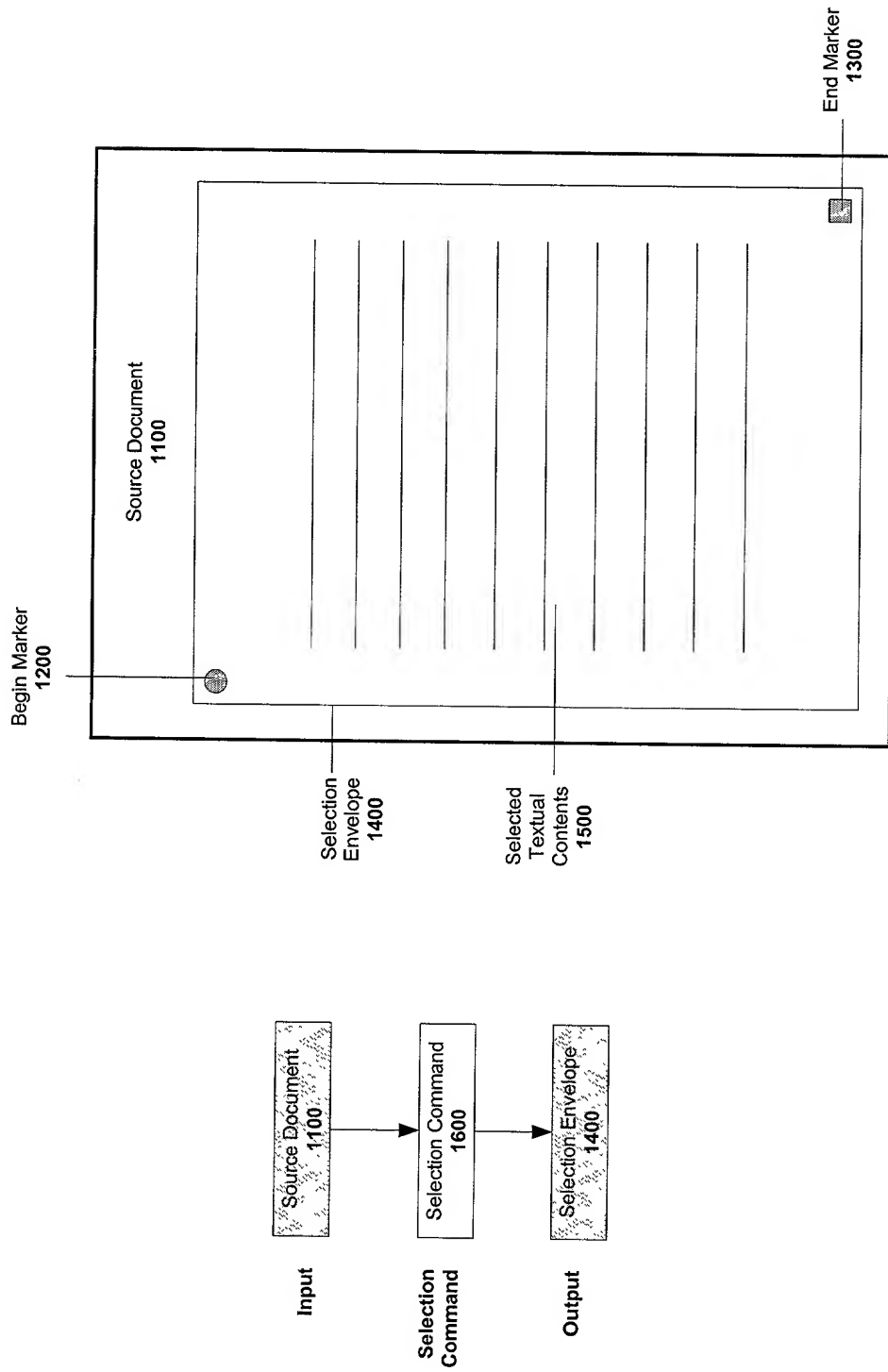


Figure 4

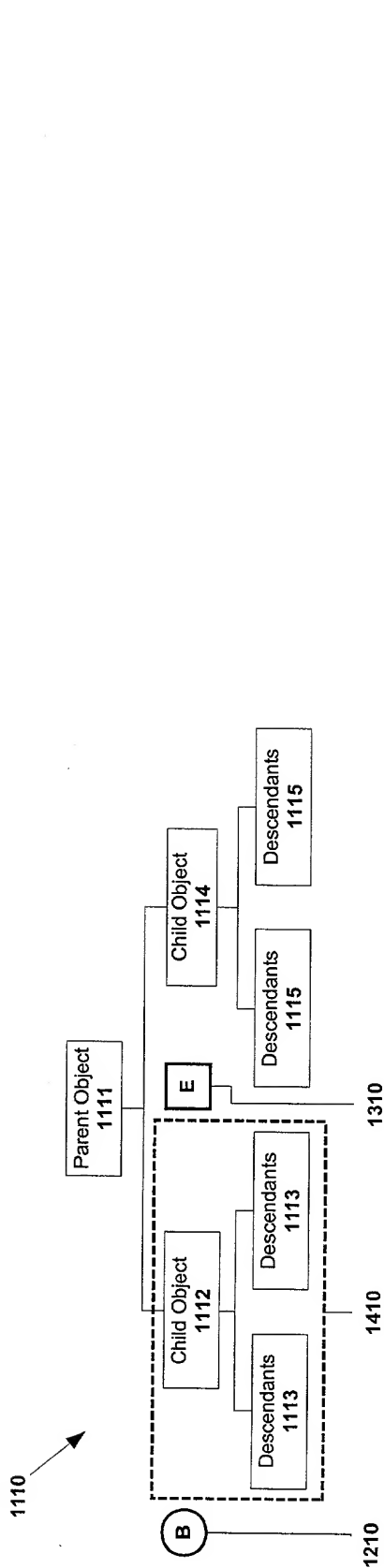


Figure 5

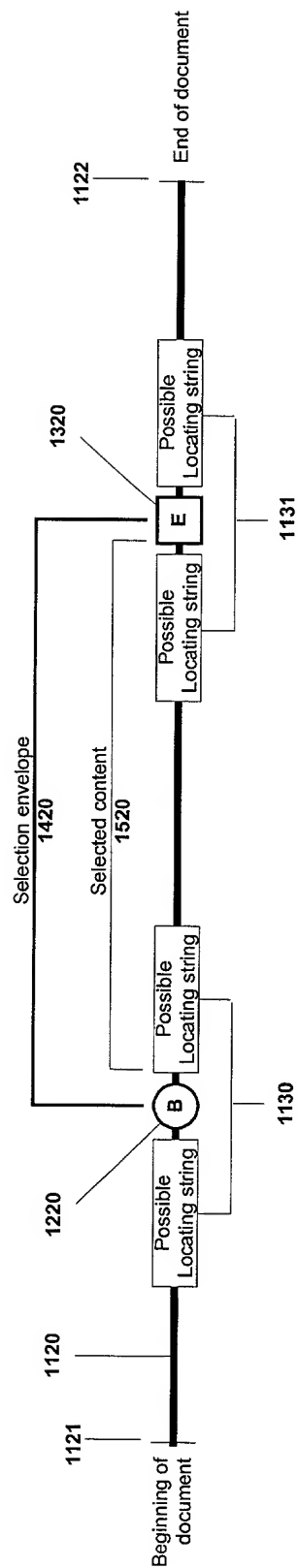


Figure 6

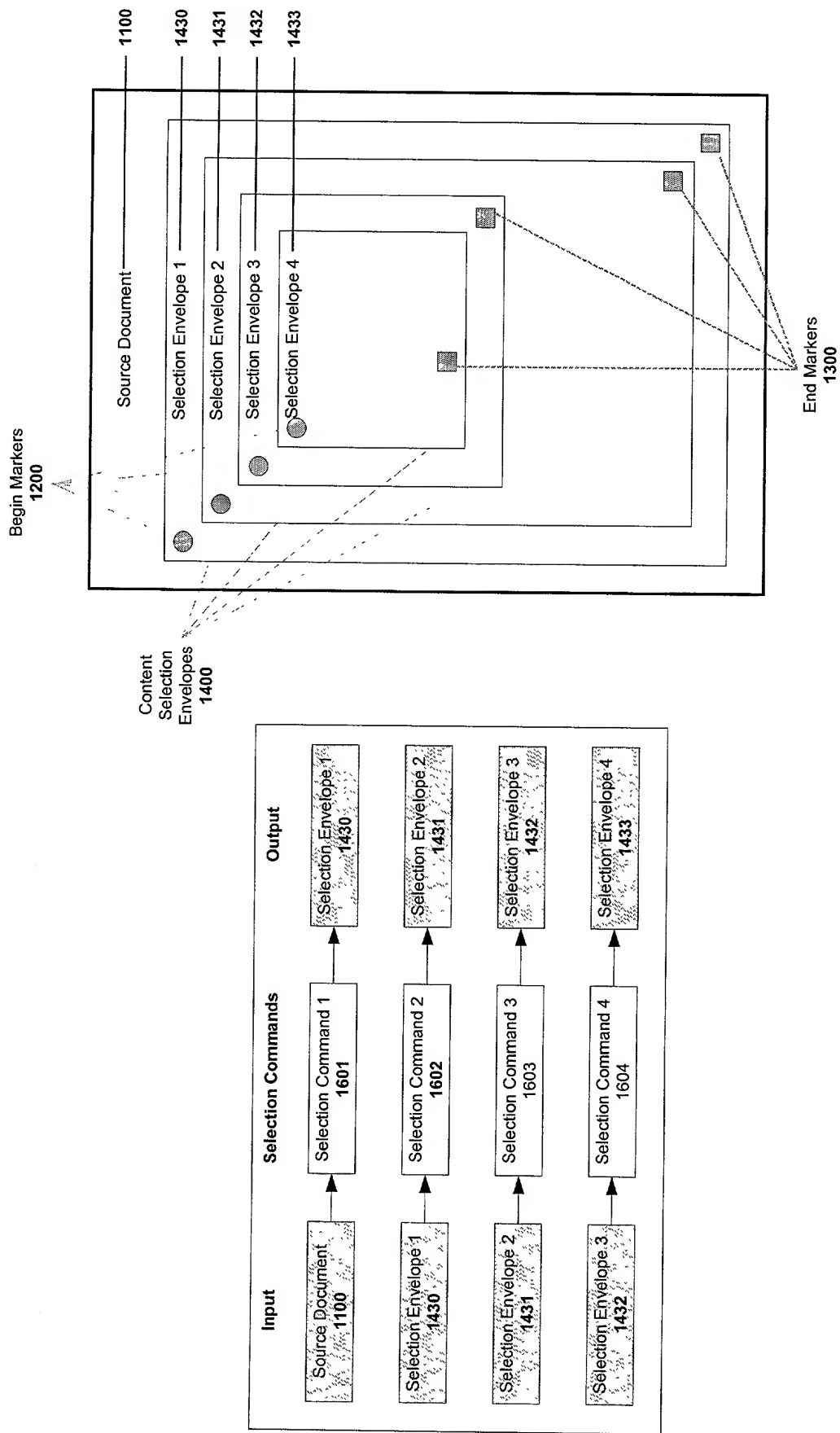


Figure 7

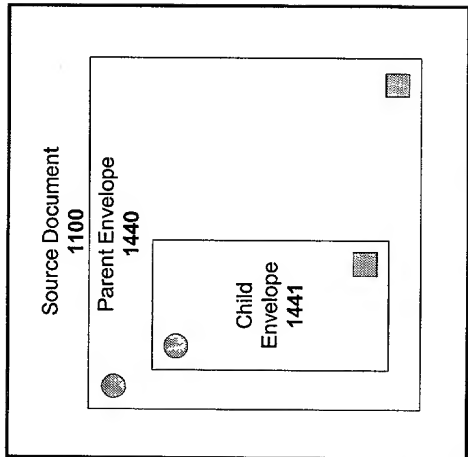


Figure 8

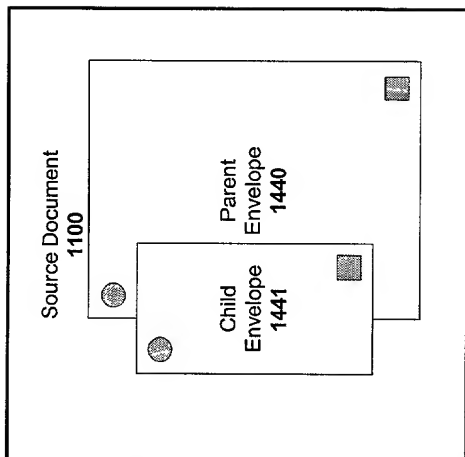


Figure 9

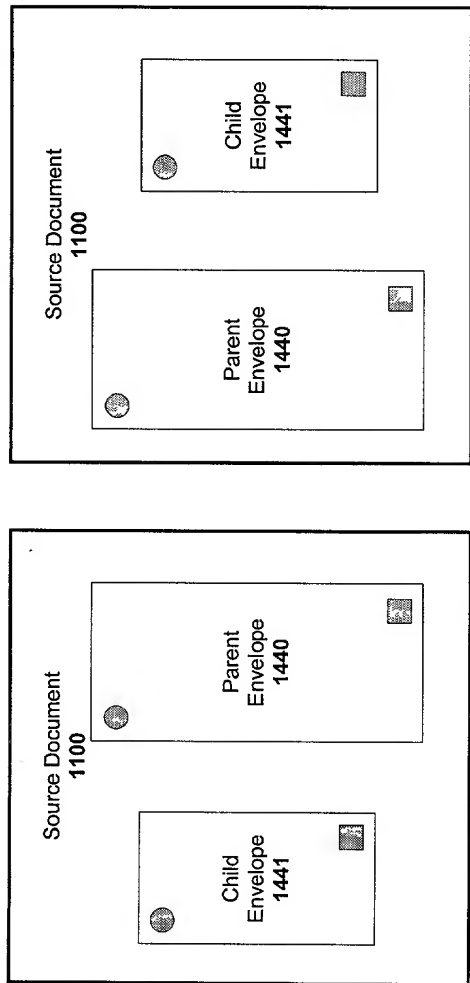


Figure 10

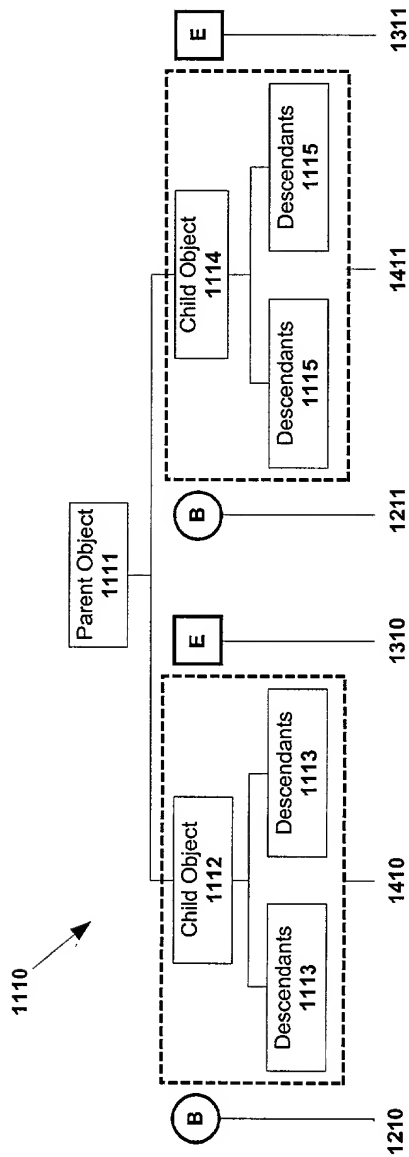


Figure 11

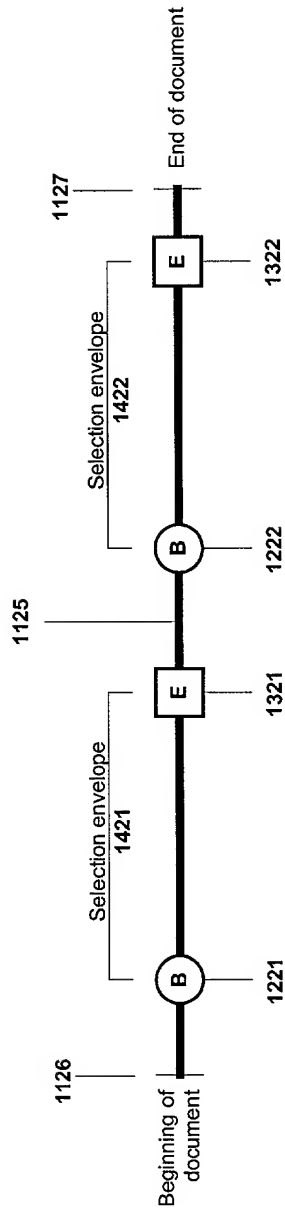


Figure 12

2000

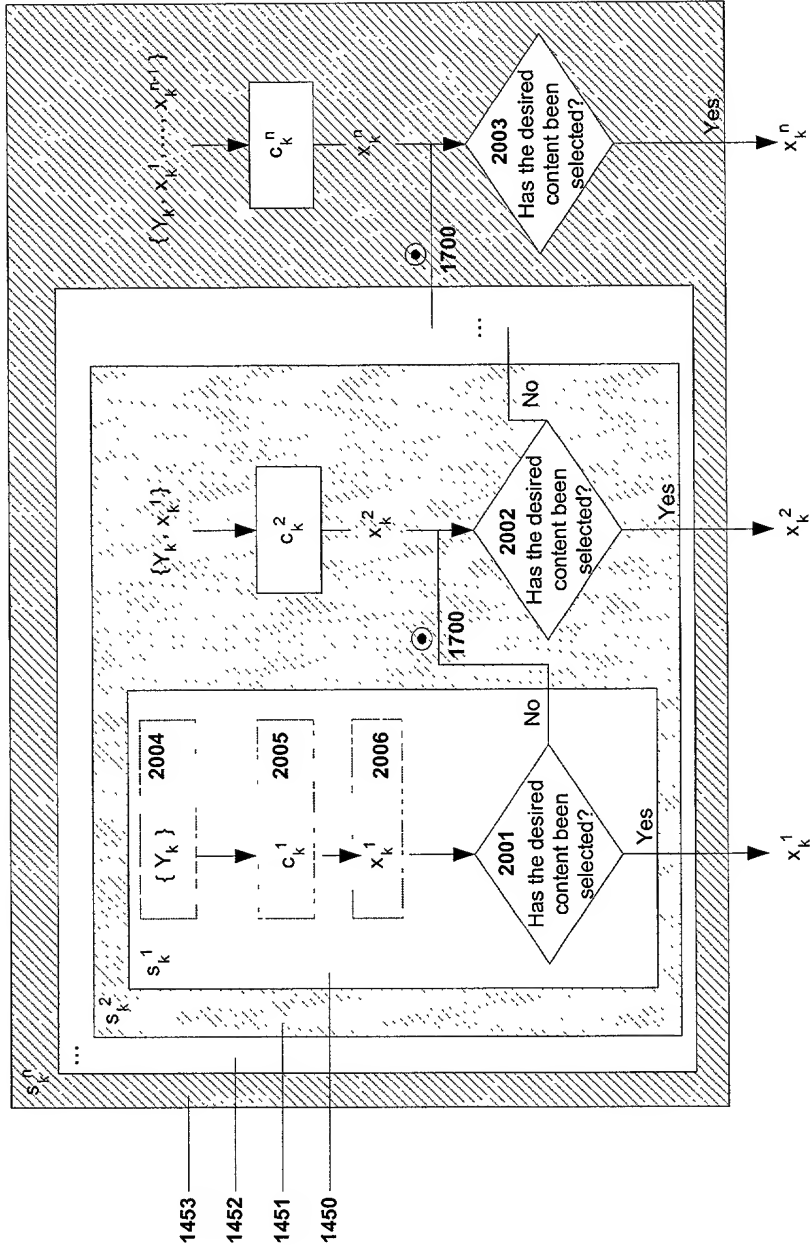
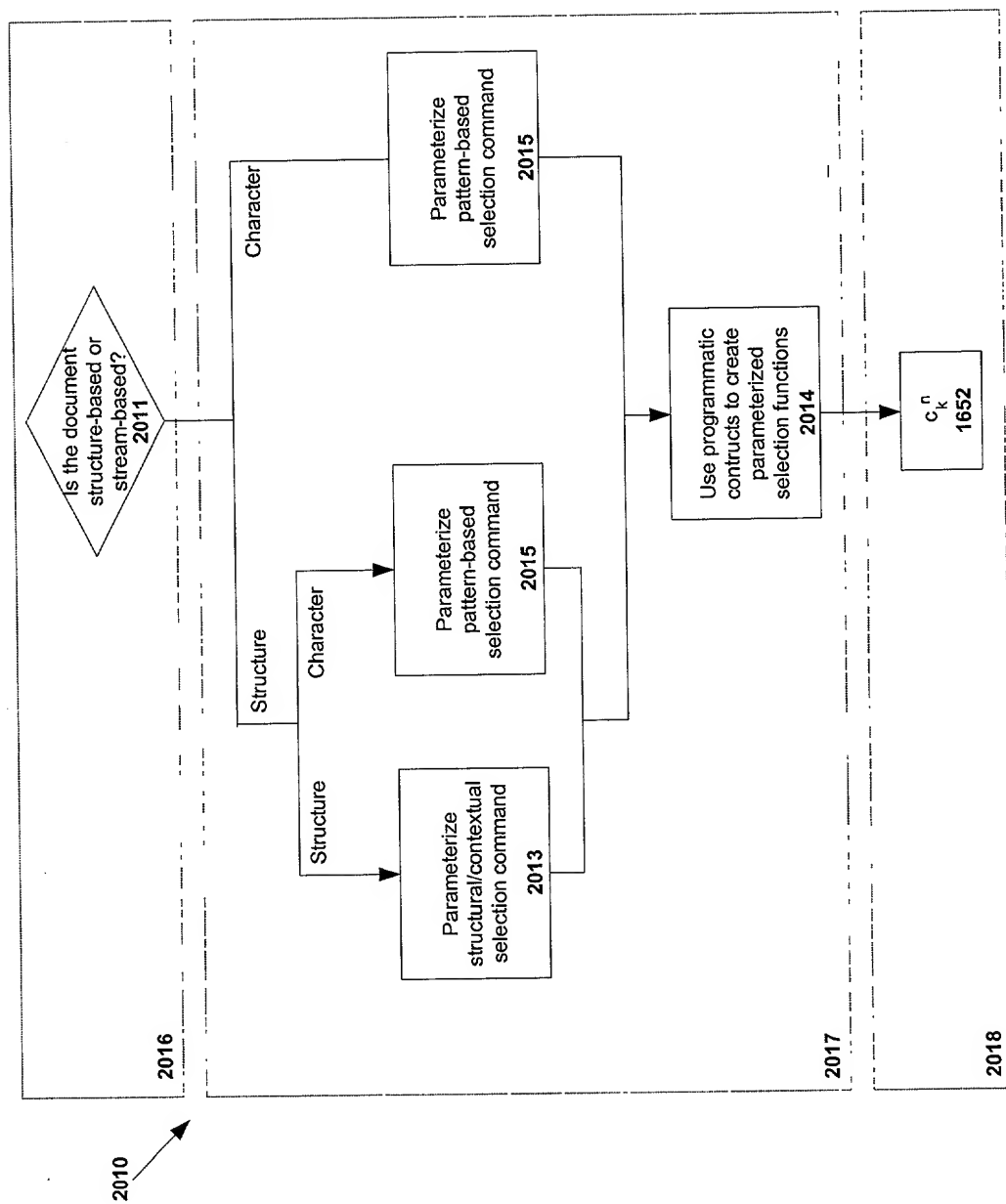


Figure 13A



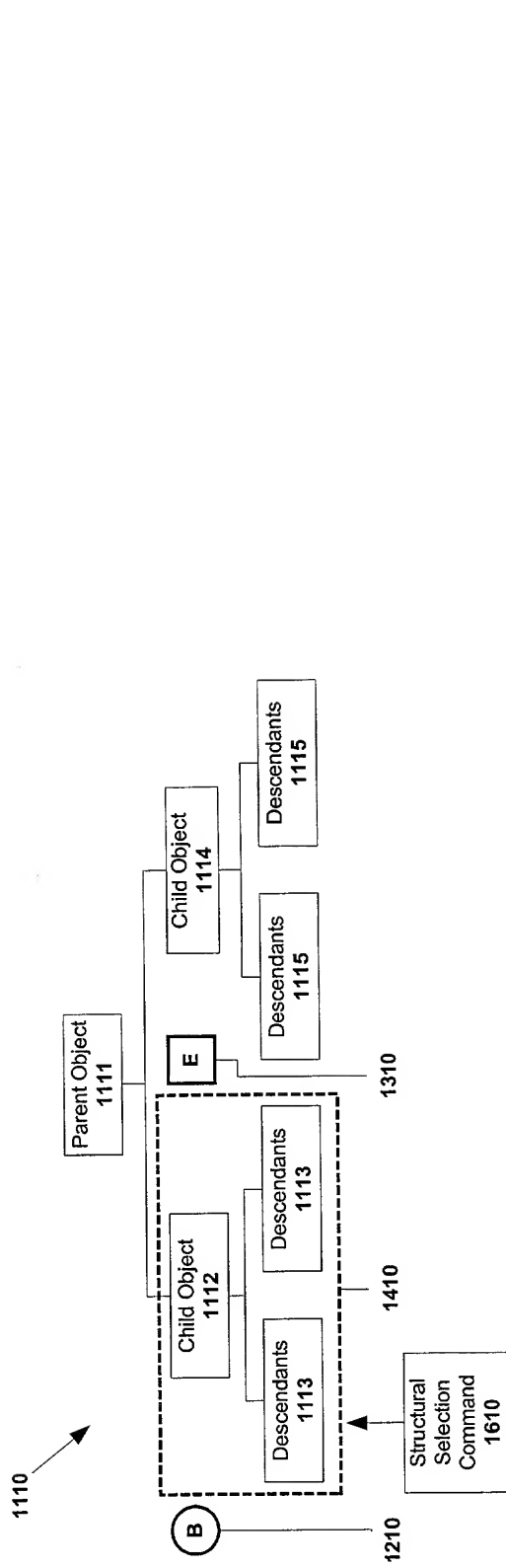


Figure 14

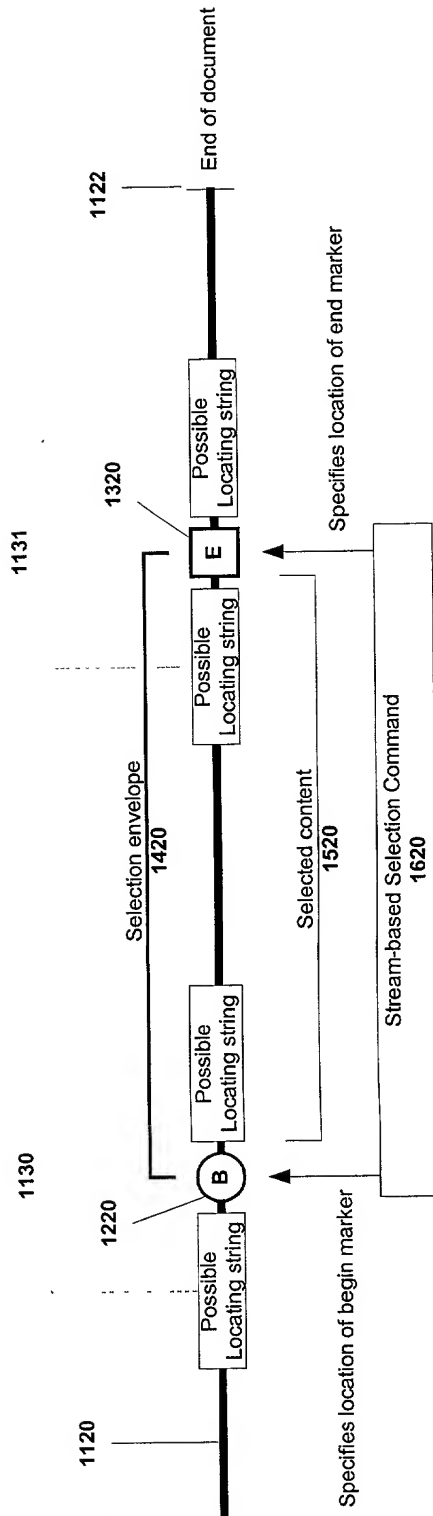


Figure 15

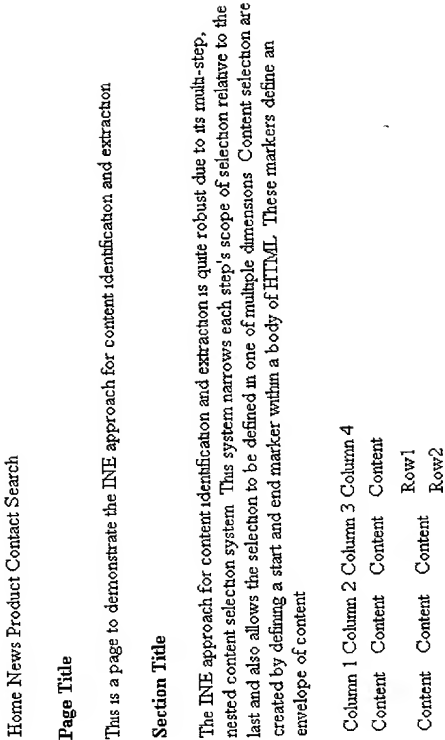


Figure 16

3013

DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.0 Transitional//EN">

<html>

```
<head>
```

<title>Sample Page</title>

</head>

<body>

<P>

<P><STRC
<TABLE>

<TR>

<TD>Home</TD>

<TD>News</TD>

 <TD>Product</TD> | <TD>Contact</TD> |

Search	
Page 1 of 1	

Page Title

<P>This is a page to demonstrate the INE approach for content identification and

extraction </P>
<P>Section Title</P>

```
<P><STRONG>Section title</STRONG></P>
<P>The INE approach for content identification and extraction is quite robust
due to its multi-step, nested content&nbps;selection system. This
system&nbps;narrows each step's scope of selection relative to the last and also
allows&nbps;the selection to be defined in one of multiple
dimensions.&nbps;Content selection are created by defining a start and end
marker within a body of HTML. These markers define an envelope of content
table>
```

<TABLE>

<TR>

	<TD>Column 1</TD>
--	-------------------

 <TD>Column 2</TD> | <TD>Column 3</TD> | <TD>Column 4</TD></TR> || <TR> | |

<TD>Content</TD>

<TD>Content</TD>	<TD>Content</TD>
------------------	------------------

<TD>Content</TD>
<TD>Content</TD>

<TD>Content</TD></TR>

<TR>
<TD>

<TD>Content</TD>
<TD>Content</TD>

<TD>Content</TD>
<TD>Content</TD>

<TD>Content</TD>
<TD>

<TD>
<TABLE>

<TABLE>

| <TR> |
 <TD>Row1</TD></TR> |

v

 <TD>Row2</TD></TR></TABLE></TD></TABLE> |

</P> </P>

<P> </P>

</body>

</body>
</html>

Figure 18

```

<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.0 Transitional//EN">
<html>
<head>
  <title>Sample Page</title>
</head>
<body>
<p><STRONG>
<TABLE>

<TR>
  <TD>Home</TD>
<TD>News</TD>
<TD>Product</TD>
<TD>Contact</TD>
<TD>Search</TD><TR></TABLE> </STRONG></P>
<p><STRONG>Page Title</STRONG></P>
<p>This is a page to demonstrate the INE approach for content identification and
extraction </P>
<p><STRONG>Section Title</STRONG></P>
<p>The INE approach for content identification and extraction is quite robust
due to its multi-step, nested content&nbsp;selection system. This
system&nbsp;narrows each step's scope of selection relative to the last and also
allows&nbsp;the selection to be defined in one of multiple
dimensions.&nbsp;Content selection are created by defining a start and end
marker within a body of HTML. These markers define an envelope of content. </P>
<TABLE>

<TR>
  <TD>Column 1</TD>
  <TD>Column 2</TD>
  <TD>Column 3</TD>
  <TD>Column 4</TD></TR>
<TR>
  <TD>Content</TD>
  <TD>Content</TD>
  <TD>Content</TD>
  <TD>Content</TD></TR>
<TR>
  <TD>Content</TD>
  <TD>Content</TD>
  <TD>Content</TD>
  <TD>Content</TD></TR>
  <TD>
    <TABLE>
      <TR>
        <TD>Row1</TD></TR>
      <TR>
        <TD>Row2</TD></TR></TABLE></TD></TR></TABLE>
<p>&nbsp;</P>
<p>&nbsp;</P>

</body>
</html>

```

Figure 19

```

<!DOCTYPE HTML PUBLIC "-//W3C/DTD HTML 4.0 Transitional/EN">
<html>
<head>
<title>Sample Page</title>
</head>
<body>
<P><STRONG>
<TABLE>

<TR>
<TD>Home</TD>
<TD>News</TD>
<TD>Product</TD>
<TD>Contact</TD>
<TD>Search</TD><TR><TABLE> <STRONG></P>
<P><STRONG>Page Title</STRONG></P>
<P>This is a page to demonstrate the INE approach for content identification and
extraction </P>
<P><STRONG> Section Title</STRONG></P>
<P>The INE approach for content identification and extraction is quite robust
due to its multi-step, nested content&nbsp;selection system. This
system&nbsp;narrows each step's scope of selection relative to the last and also
allows&nbsp;the selection to be defined in one of multiple
dimensions&nbsp;. Content selection are created by defining a start and end
marker within a body of HTML. These markers define an envelope of content
<TABLE>

<TR>
<TD>Column 1</TD>
<TD>Column 2</TD>
<TD>Column 3</TD>
<TD>Column 4</TD></TR>
<TR>
<TD>Content</TD>
<TD>Content</TD>
<TD>Content</TD>
<TD>Content</TD></TR>
<TR>
<TD>Content</TD>
<TD>Content</TD>
<TD>Content</TD>
<TD>Content</TD>
<TD>
<TABLE>
<TR>
<TD>Row1</TD></TR>
<TR>
<TD>Row2</TD><TR><TABLE><TD></TR></TABLE>
<P>&nbsp;</P>
<P>&nbsp;</P>
</body>
</html>

```

3001

3002

Figure 20


```
<html>
<head>
<title>Sample Page</title>
```

```
<body>
<P><STRONG>
<TABLE>
```

| <TR> |

<TD>News</TD>	<TD>Product</TD>	<TD>Contact</TD>
---------------	------------------	------------------

<TD>Contact</TD>
<TD>Search</TD></TR></TABLE> </P>

<|ID>Search</ID></TABLE></P><P>Page Title</P>

This is a page to demonstrate the INE approach for content identification and extraction

3001	extraction </P> <P>
------	--------------------------------

3003

4>The INE approach for content identification and extraction is quite robust due to its multi-step, nested content selection system. This system narrows each step's scope of selection relative to the last and also allows the selection to be defined in one of multiple dimensions .Content selection is created by defining a start and end marker within a body of HTML. These markers define an envelope of content <TABLE>

```
<TR>  
<TD>Column 1</TD>  
<TD>Column 2</TD>  
<TD>Column 3</TD>  
<TD>Column 4</TD></TR>  
  
<TR>  
<TD>Content</TD>  
<TD>Content</TD>  
<TD>Content</TD>  
<TD>Content</TD></TR>  
  
<TR>  
<TD>Content</TD>  
<TD>Content</TD>  
<TD>Content</TD>  
<TD></TD>  
</TR>
```

```
<TR>  
  <TD>Row1</TD></TR>  
<TR>  
  <TD>Row2</TD></TR></TABLE></TD></TR></TABLE>  
  
<p>&nbspbsp:</P>  
<p>&nbspbsp:</P>
```

3002

Figure 21

<DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.0 Transitional//EN">

<html>
<head>
 <title>Sample Page</title>

</head>

<body>
<p>
<TABLE>

<TR>
<TD>Home</TD>
<TD>News</TD>
<TD>Product</TD>
<TD>Contact</TD>
</TR>
<TR>Search</TR>
</TABLE>
<p>Page Title<p>
<p>This is a page to demonstrate the INE approach for content identification and extraction.</p>
<p>This is a new paragraph</p>

<HR>
<TABLE>

<TR>
<TD>Section Title</TD></TR>
</TABLE>
<p>The INE approach for content identification and extraction is quite robust due to its multi-step, nested content selection system. This system narrows each step's scope of selection relative to the last and also allows the selection to be defined in one of multiple dimensions . Content selection is created by defining a start and end marker within a body of HTML. These markers define an envelope of content.</p>
<TABLE>
<TR>
<TD>Column 1</TD>
<TD>Column 2</TD>
<TD>Column 3</TD>
<TD>Column 4</TD></TR>
</TR>
<TD>Content</TD>
<TD>Content</TD>
<TD>Content</TD>
<TD>Content</TD>
</TR>
<TD>Content</TD>
<TD>Content</TD>
<TD>Content</TD>
<TD>
</TABLE>

<TR>
<TD>Row1</TD></TR>
<TR>
<TD>Row2</TD></TR>
</TABLE>
<p> </p>
<p> </p>
</body>
</html>

Figure 22

```

<html>
<head>
<title>Sample Page</title>
</head>
<body>
<P><STRONG>
<TABLE>
<TR>
<TD>Home</TD>
<TD>News</TD>
<TD>Product</TD>
<TD>Contact</TD>
</TR>
<TR>
<TD>Search</TD><TR></TABLE> <STRONG><P>
<P><STRONG>Page Title</STRONG><P>
extraction.</P>
<P><STRONG>Section Title</STRONG><P>
<P> The INE approach for content identification and extraction is quite robust
due to its multi-step, nested content&nbsp;&nbsp;&nbsp;selection system. This
system&nbsp;&nbsp;&nbsp;narrows each step's scope of selection relative to the last and also
allows&nbsp;&nbsp;&nbsp;the selection to be defined in one of multiple
dimensions.&nbsp;&nbsp;&nbsp;Content selection are created by defining a start and end
marker within a body of HTML. These markers define an envelope of content
</TABLE>
<TR>
<TD>Column 1</TD>
<TD>Column 2</TD>
<TD>Column 3</TD>
<TD>Column 4</TD></TR>
<TR>
<TD>Content</TD>
<TD>Content</TD>
<TD>Content</TD>
<TD>Content</TD>
</TR>
<TR>
<TD>Content</TD>
<TD>Content</TD>
<TD>Content</TD>
<TD>Content</TD>
</TR>
<TABLE>
<TR>
<TD>Row1</TD></TR>
<TR>
<TD>Row2</TD></TR></TABLE></TD></TR></TABLE>
<P>&nbsp;&nbsp;&nbsp;</P>
<P>&nbsp;&nbsp;&nbsp;</P>
</body>
</html>

```

Figure 23

January 16, 2002¶

Arthur Andersen Fires an Executive for Enron Orders¶

By RICHARD A. OPPEL Jr. with KURT EICHENWALD¶

WASHINGTON, Jan. 15 — Arthur Andersen fired its partner in charge of auditing the Enron Corporation today, saying he had ordered the destruction of thousands of documents and e-mail messages after learning that the Securities and Exchange Commission had begun an investigation of Enron's accounting. ¶

The fired partner, David B. Duncan, called a meeting of auditors at the firm's Houston office and ordered "an expedited effort to destroy documents" on Oct. 23, the day after Enron disclosed that the S.E.C. had begun its inquiry, the firm said. The destruction apparently did not end until Mr. Duncan's assistant sent an e-mail message to other secretaries on Nov. 9 that said "stop the shredding," the firm said. Andersen had received a subpoena from the S.E.C. the day before. ¶

Further disclosures today underscored the accountants' central role in the scandal surrounding Enron, the pioneering Houston energy company that was the nation's seventh-largest company before its precipitous fall into bankruptcy protection. ¶

A report in October by Enron's law firm, commissioned by the company to investigate an employee's allegations of improper accounting, concluded that Andersen auditors reviewed and approved of transactions by Enron-related partnerships that contributed to the company's collapse. The employee, who made the allegations in an anonymous letter, has been identified as Sherron S. Watkins, an Enron vice president for corporate development. ¶

Figure 24

4001 4002

January 16, 2002¶

Arthur Andersen Fires an Executive for Enron Orders¶
By RICHARD A. OPPEL Jr. with KURT EICHENWALD¶

4000

WASHINGTON, Jan. 15 — Arthur Andersen fired its partner in charge of auditing the Enron Corporation today, saying he had ordered the destruction of thousands of documents and e-mail messages after learning that the Securities and Exchange Commission had begun an investigation of Enron's accounting. ¶

The fired partner, David B. Duncan, called a meeting of auditors at the firm's Houston office and ordered "an expedited effort to destroy documents" on Oct. 23, the day after Enron disclosed that the S.E.C. had begun its inquiry, the firm said. The destruction apparently did not end until Mr. Duncan's assistant sent an e-mail message to other secretaries on Nov. 9 that said "stop the shredding," the firm said. Andersen had received a subpoena from the S.E.C. the day before. ¶

Further disclosures today underscored the accountants' central role in the scandal surrounding Enron, the pioneering Houston energy company that was the nation's seventh-largest company before its precipitous fall into bankruptcy protection. ¶

4003

A report in October by Enron's law firm, commissioned by the company to investigate an employee's allegations of improper accounting, concluded that Andersen auditors reviewed and approved of transactions by Enron-related partnerships that contributed to the company's collapse. The employee, who made the allegations in an anonymous letter, has been identified as Sherron S. Watkins, an Enron vice president for corporate development. ¶

Figure 25